DOCUMENT RESUME

ED 421 527 TM 028 853

AUTHOR Bridgeman, Brent; Potenza, Maria
TITLE Effects of an On-Screen versus Bring-Your-Own Calculator Policy on Performance on the Computerized SAT I: Reasoning Test in Mathematics. .
PUB DATE 1998-04-00
NOTE 19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Diego, April 12-16, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Ability; *Calculators; College Entrance Examinations; *Computer Assisted Testing; Ethnic Groups; *High School Students; High Schools; *Mathematics Tests; *Online Systems; Student Attitudes
IDENTIFIERS *Scholastic Assessment Tests

ABSTRACT

Students taking the paper-based Scholastic Assessment Test (SAT) mathematics test are permitted to bring and use their own hand-held calculators, and this policy was continued for the computer-adaptive tests (CAT) designed for use in talent search programs. An on-screen calculator may also be used with the CAT. The bring-your-own option has raised some fairness concerns (because all students cannot afford the most sophisticated calculators) as well as security concerns (because questions could be entered into a calculator's memory and taken from the testing session), but forcing all students to use an unfamiliar on-screen calculator raises different fairness issues. This study of the computerized SAT compared the performance of 360 students tested under the current policy (bring-your-own or on-screen) with the performance of 373 students who had only an on-screen calculator available. These students had already taken a national administration of the SAT and were invited to participate. Across ethnic, gender, and ability groups, students who had to use the on-screen calculator performed as well as students who were permitted to use their own calculators. Nevertheless, students expressed a strong preference for using their own calculators. (Contains eight tables and one figure.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 421 527

Effects of an On-Screen Versus Bring-Your-Own Calculator Policy On Performance on the Computerized SAT I: Reasoning Test in Mathematics

Brent Bridgeman
and
Maria Potenza

Brent Bridgeman

Educational Testing Service
Princeton, NJ 08541

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
✓ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM028853

Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, April 1998.

Abstract

Students taking the paper-based SAT-M are permitted to bring and use their own hand-held calculators, and this policy was continued for the computer-adaptive test (CAT) designed for use in talent search programs. An on-screen calculator may also be used with the CAT. The "bring-your-own" option has raised some fairness concerns (because all students cannot afford the most sophisticated calculators) as well as security concerns (because questions could be entered into a calculator's memory and taken from the testing session); forcing all students to use an unfamiliar on-screen calculator raises different fairness issues. This study of the computerized SAT compared the performance of 360 students tested under the current policy (bring-your-own or on-screen) with the performance of 373 students who had only an on-screen calculator available. Across ethnic, gender, and ability groups, students who had to use the on-screen calculator performed as well as students who were permitted to use their own calculators. Nevertheless, students expressed a strong preference for using their own calculators.

Effects of an On-Screen Versus Bring-Your-Own Calculator Policy On Performance on the Computerized SAT I: Reasoning Test in Mathematics

The SAT I: Reasoning Test is currently offered as a computer-adaptive test (CAT) for students in talent search programs, and in the future this SAT CAT may be offered to a much wider audience. For the mathematics portion of this computer-delivered test, students may use either their own calculators that they bring with them to the testing center or an on-screen calculator. Other computerized tests, such as the academic skills assessments that are part of The Praxis Series: Professional Assessments for Beginning Teachers, offer only an on-screen calculator. Permitting students to bring their own calculators has raised security concerns because questions could be entered into a calculator's memory and taken away from the testing session. Furthermore, the bring-your-own policy may create a perception of unfairness as some students bring simple 4-function calculators while others bring sophisticated and relatively expensive graphing calculators. On the other hand, being forced to use an unfamiliar on-screen calculator may create problems for students who have spent months or years becoming accustomed to their own calculators.

Because the mathematics portion of the SAT (SAT-M) emphasizes reasoning skills, and not computational facility, no complex computations are required and a calculator is not needed. Nevertheless, previous research with the SAT has shown that students who have access to their own calculators get higher scores, on average, than students who do not have any access to a calculator (Bridgeman, Harvey, & Braswell, 1995). This study was designed to determine whether students who may bring their own calculators are advantaged relative to students who must use the on-screen calculator, and to better understand the effects of calculator usage in order to provide better advice to students.

Method

Sample and Procedures

A sample of students who took the regular administration of the May SAT I: Reasoning Test, who lived within 15 miles of one of the designated testing centers, and who did not repeat the test in June (in either a standard or special CAT administration), were invited to participate in the calculator study. The incentive to participate was the opportunity to see their scores immediately and to decide whether they wanted them to be added to their permanent score record (which is sent to the colleges of their choice) or canceled.

Invitations were sent to a stratified sample of 3000 students in the hope of obtaining a final sample of at least 600 students. The sample was stratified into a high-ability group and a low-ability group based on SAT-M scores from the May national administration of the SAT; high-ability was defined as those students with SAT-M scores of 550 or higher (Level 2) and low ability was defined as 450 or lower (Level 1). Within each of these strata, half of the invitees were told to bring their own calculators while the other half were told that they could use only an on-screen calculator. We had intended for these groups within strata to be randomly selected, but, through a misunderstanding with the people mailing the letters, selection was not strictly random. Rather, invitees were selected from a list of registration numbers from the May test. The registration numbers had been assigned in the order that requests to take the May test were received. Within each ability stratum, the first 750 registration numbers for students meeting the eligibility criteria were assigned to the Bring-Your-Own calculator group and the next 750 were assigned to the On-Screen group. The impact of this selection procedure is discussed in the Results and Discussion section.

The group that was allowed to bring their own calculators was told in the invitation letter, "Although you may use the on-screen calculator, you are strongly encouraged to bring a calculator that you are comfortable using." The invitation explained that not all volunteers would be able to be accommodated but that reservations would be accepted on a first-call first-served basis. The letter further indicated that two types of on-screen calculators (four-function and scientific) would be available. At the beginning of the test, the students would decide which type of on-screen calculator they wanted to use and only this calculator would be available for the test. The invited students were also sent a *Bulletin for the Computerized SAT Field Trial* that explained the administration details for the Computerized SAT and how the on-screen calculators worked.

As students were checked-in to the testing center, the administrator was to verify that students in the on-screen group did not bring calculators. After the administrator helped the student select either the four-function or scientific on-screen calculator, the students completed the rest of the testing session on their own. Students first completed tutorials that explained the mechanics of testing including how to use a mouse, how to scroll, how to answer test questions, and how to use the on-screen calculator. Students could spend as long as they wanted in this tutorial session. At the end of the tutorials, six questions on general computer familiarity (e.g., how often do you use a computer at home) were presented. Next, students responded to the 33 verbal questions (in 60 minutes) followed by a ten minute break and then the 28 mathematics questions (in 60 minutes). Finally, the students responded to 18 questions in an on-screen posttest questionnaire.

Results and Discussion

Final Analysis Sample

Because the budget could not support testing all students who might volunteer, we stopped accepting reservations two weeks after the invitation letters were sent out and 912 reservations already had been received. Out of these 912 volunteers, 774 (84%) kept their testing appointments. Six cases were lost when their scores could not be matched to their May scores. Preliminary analyses of the posttest questionnaire responses of these 768 students suggested that there may have been a few lapses in monitoring calculator use among the students who were not supposed to use their own calculators. Specifically, 32 students in this group reported in the posttest questionnaire (question 17) that they had used their own calculators either exclusively (26 students) or that they had used both their own calculators and the on-screen calculator (6 students). Although a few students may have inadvertently checked the wrong choice in the questionnaire, other indicators suggested that most of these students had indeed used their own calculators. Thus, on question 11, only one of the 32 students indicated that they were not permitted to use their own calculators. Furthermore, all 32 indicated that they had used a calculator on more than one or two questions, but an actual count of on-screen calculator use from the computer's internal record indicated that 23 of these students never used the on-screen calculator and 5 more used the on-screen calculator only one or two times. Therefore, these 32 students (plus three more who did not answer the relevant questionnaire question) were removed from further analyses, and the final sample contained 733 students whose questionnaire responsed were consistent with the group to which they were assigned.

Impact of Assignment Procedure

The impact of assigning students to groups by registration number from the May SAT was investigated by comparing the groups within strata on a number of variables that might be related to performance on a mathematics test. These results are summarized in Table 1. Except for the May SAT-M scores, these scores are self reports from the Student Descriptive Questionnaire (SDQ) that students completed when they registered to take the May test. High school grade point average (GPA) was reported on a 12-point scale from A+ through F (with no D- or F+), letter grades were transformed to a numerical scale (A+ =4.3, A = 4.0, A- = 3.7, B+ = 3.3 ..F = 0.0). The SDQ asks students to report the number of years in high school that they had studied or planned to study various courses; "years advanced math/science" is the sum of the years reported for trigonometry, precalculus, calculus, chemistry, and physics. The *n*s vary slightly because not all students respond to all questions on the SDQ.

For both levels and both genders, mean scores on the May SAT-M were higher for the On-Screen group than for the Bring-Your-Own group. Although these differences were not large relative to the size of the within-group standard deviations, the condition main effect was statistically significant in a 2 conditions x 2 levels x 2 genders ANOVA (F [1,725] = 4.57, p = .03); condition did not significantly interact with either level or gender. Within Level 1, males were somewhat overrepresented in the On-Screen group, and females were overrepresented in this group in Level 2. For the high school GPA and years of math/science, differences were generally in the same direction, but were quite small in absolute terms and did not reach statistical significance at the .05 level. Thus, the observable bias in the sample, though real, did not appear to be large enough to seriously bias the results. Although there is no totally adequate solution for

Table 1
Comparison of Groups on Pretest Variables

		Level 1				Level 2			
		M		F		M		F	
Variable	Statistic	Own	On-Screen	Own	On-Screen	Own	On-Screen	Own	On-Screen
May	*n*	47	66	102	91	123	99	88	117
SAT-M	M	397	410	402	407	625	634	609	616
	SD	48	39	42	33	64	61	46	52
High	*n*	45	61	95	87	16	96	83	111
School	M	2.78	2.82	2.96	3.04	3.62	3.60	3.66	3.81
GPA	SD	.52	.61	.57	.48	.53	.50	.47	.36
Years	*n*	39	57	96	89	115	98	85	111
Advanced	M	1.36	1.55	1.63	1.89	3.64	3.56	3.72	3.74
Math/Science	SD	1.1	1.1	1.1	1.4	1.4	1.7	1.3	1.3
Ethnic Group	%White	72	76	72	73	77	70	73	71
	% African American	26	12	16	18	2	2	2	1
	% Asian American	2	3	5	5	15	16	15	16

a breakdown in random assignment, we decided to mute its impact by using the May SAT-M scores as a covariate and using gender as an independent variable.

Effects of Calculator Type (On-Screen versus Bring-Your-Own)

The 2 (types) x 2 (levels) x 2 (genders) analysis of covariance, with SAT-M scores on the computerized SAT as the dependent variable, indicated no significant difference between the On-Screen and Bring-Your-Own groups (F [1,724] = 1.29, p = .26), and no significant interactions with the calculator-type condition. The adjusted means in the two calculator conditions differed by only four points. Thus, for both men and women of high and low mathematical ability, being forced to use an on-screen calculator should not have a negative impact on test performance. Although sample sizes were small for exploratory analyses of covariance within the three minority groups (African American, Asian American, and Latino), no significant differences were found with all Fs for the condition effect less than 1.0.

A number of additional ANCOVAs were run for groups defined by various responses on the posttest questionnaire. Because these analyses were strictly exploratory, we made no attempt to adjust the experimentwise error rate, using the .05 level in each analysis. There were no significant calculator-type effects for the following groups: 432 students who reported that they used a calculator on every or almost math test, 262 students who reported that they routinely used a graphing calculator, 221 students who reported that they had taken (or would take) at least one year of calculus in high school, and 183 students who reported using the calculator's trigonometric functions on the test. The one significant difference found was for the 160 students (89 in the Bring-Your-Own group and 71 in the On-Screen group) who reported that they used the calculator on more than half of the questions on the test. For this group, the adjusted mean of 539 in the Bring-Your-Own group was significantly higher than the mean of 515 in the On-Screen

group, F (1, 151) = 9.90, p = .002, with no significant interactions with level or gender. Thus, students who used a calculator frequently on the test seemed to do better with their own calculators. A possible reason for this is that it takes more time to use the on-screen calculator. Because the computerized SAT is not very speeded, this extra time would not hurt most users, but it could be a problem for students who use a calculator on a substantial number of the questions. We confirmed that the On-Screen group indeed took more time by analyzing the total number of seconds it took each group to complete the mathematics section. For this analysis, we used the total time for all other sections of the test as a covariate. The On-Screen group was significantly slower (F [1, 724] = 6.49, p = .01) with an adjusted mean of 2796 seconds compared to 2699 in the Bring-Your-Own group.

Questionnaire Results

Preference. Although performance may be equally good with an on-screen calculator or the student's own calculator, the student may still prefer one type of calculator over the other. As shown in Table 2, responses to questionnaire item 18 (Q18) indicated that a majority of students in all groups would prefer to use their own calculators. For the students who were allowed to use only the on-screen calculator, this preference was also clearly seen in the responses to Q16, as shown in Table 3. Although students at both ability levels expressed a desire to be able to use their own calculators, this preference was especially strong among the more able students.

Table 2

Q18--On a computer test such as the one you just took, would you prefer to use the on-screen calculator or your calculator?

	Level 1			Level 2		
Responses	Bring Own	On-Screen	Total	Bring Own	On-Screen	Total
Would not use either	0	1	1	1	1	1
Strongly prefer on-screen	7	15	11	1	1	1
Strongly prefer mine	66	40	52	73	53	63
Slightly prefer on-screen	1	9	5	2	3	3
Slightly prefer mine	10	17	14	12	31	21
No preference	16	19	17	11	12	11

Note --Entries are the percentage of students within each category (column).

Table 3

Q16--While using the computer's on-screen calculator, did you wish you could use your own calculator? (On-screen group only)

Responses	Level 1	Level 2
Did not use on-screen calculator	1	1
No	39	17
Yes, occasionally	41	40
Yes, frequently	18	43

Note --Entries are the percentage of students within each category (column).

Questionnaire item 15 was designed to determine whether a preference for their own calculators might be related to the on-screen calculator blocking the view of the test question while it was being used. However, as indicated in Table 4, only 8% of the level 1 students and 16% of the level 2 students found the on-screen calculator blocking the question to be annoying. Thus, changing the size or location of the on-screen calculator would probably do little to increase its popularity relative to the students' own calculators.

Table 4

Q15--If you used the computer's on-screen calculator, did it ever interfere with your ability to see the question?
(On-screen group only)

Responses	Level 1	Level 2
Did not use on-screen calculator	2	1
No	36	19
Occasionally, but not problem	46	55
Occasionally, and annoying	5	11
Frequently, but not problem	8	9
Frequently, and annoying	3	5

Calculator type. As indicated in Table 5, over 95% of the students in both levels reported using a calculator when they took the paper-based SAT I in May. Most Level 2 students used a graphing calculator while most Level 1 students used a scientific calculator. Relatively few students in either level used a 4-function calculator.

Table 5

Q10--When you took the paper-based SAT I in May, what type of calculator did you use?

	Level 1			Level 2		
Responses	Bring Own	On-Screen	Total	Bring Own	On-Screen	Total
None	5	3	4	1	1	1
4-Function	15	12	13	5	2	4
Scientific	70	64	67	43	40	42
Graphing	10	20	15	50	57	54
Other	1	1	1	0	0	0

Table 6 shows the same general pattern for the types of calculators brought to the computer-delivered test by students in the group who were allowed to bring their own calculators. The

percentage of students choosing not to bring a calculator was somewhat higher for the computer test, though still low in absolute terms, probably because students had the option of using the on-screen calculator even though they were encouraged to use their own. It is not clear whether the small percentage reporting that they were not permitted to bring a calculator reflects inattention to the question, random marking to get through the questionnaire, or possibly a parent or counselor who suggested that they should not bring a calculator.

Table 6

Q11--What type of calculator did you bring with you today?
(Bring own group only)

Responses	Level 1	Level 2
Not permitted to bring	3	4
Chose not to bring	14	9
4-Function	11	4
Scientific	55	34
Graphing	16	50
Other	1	0

Before the test began, students were asked to select the type (4-function or scientific) of on-screen calculator that would be available for use during the test. In the on-screen group, 76% of the students in Level 1 and 90% in Level 2 selected the scientific calculator.

As indicated in Table 7, the basic functions of a 4-function calculator were used by most students. The most commonly used functions were multiplication and division (87 to 94 percent across groups) with slightly less use of the calculator for addition and subtraction (83 to 85 percent across groups). Trigonometric functions were used more frequently within the higher ability group, but even in that group only 6% used these functions. Of the functions not available on the 4-function calculator, the most heavily used was one permitting raising numbers to powers

higher than 2, with this function used by over a third of the students at both ability levels who brought their own calculators. At both ability levels, this function was used more by students who could bring their own calculators than by students who used the on-screen calculator. The graphing function, which was not available in the on-screen scientific calculator, was used by 9% of the students in the high ability group, but by no one in the lower ability group.

Table 7

Q13--Which calculator functions did you use during today's Computerized SAT? MARK ALL ANSWERS THAT APPLY.

	Level 1			Level 2		
Responses	Bring Own	On-Screen	Total	Bring Own	On-Screen	Total
None	1	1	1	1	1	1
Add or subtract	83	83	83	85	83	84
Multiply	91	93	92	94	94	94
Divide	91	87	89	91	93	92
Square root	27	33	30	48	53	51[b]
Percent	38	24[a]	30	32	14[a]	23[b]
Trigonometric functions	3	1	2	5	7	6[b]
Powers greater than 2	37	20[a]	28	44	26[a]	35
"Solve" or "solver"	0	2	1	5	1	3
Graphing function	0	0	0	9	0	5[b]

[a] Within level difference between Bring Own and On-Screen groups significant at .05.
[b] Difference between totals for Levels 1 and 2 significant at .05.

Frequency of calculator use. Table 8 indicates that most students used a calculator on at least three questions with about 20 percent at both ability levels using the calculator for more than half of the questions.

Table 8

Q14--On how many questions did you use a calculator today?

	Level 1			Level 2		
Responses	Bring Own	On-Screen	Total	Bring Own	On-Screen	Total
None	1	1	1	1	1	1
One or two	5	6	5	6	5	5
More than two, < half	70	73	71	68	77	73
More than half	25	20	22	26	18	22

For students who used the on-screen calculator we did not have to rely on self-reports because the system automatically recorded, for each question, whether or not the calculator was used. As indicated in Figures 1 and 2, there was a broad distribution of calculator use with average calculator use of about 9-10 questions at both ability levels. Thus, on average, students used a calculator on about one third of the 28 question test.

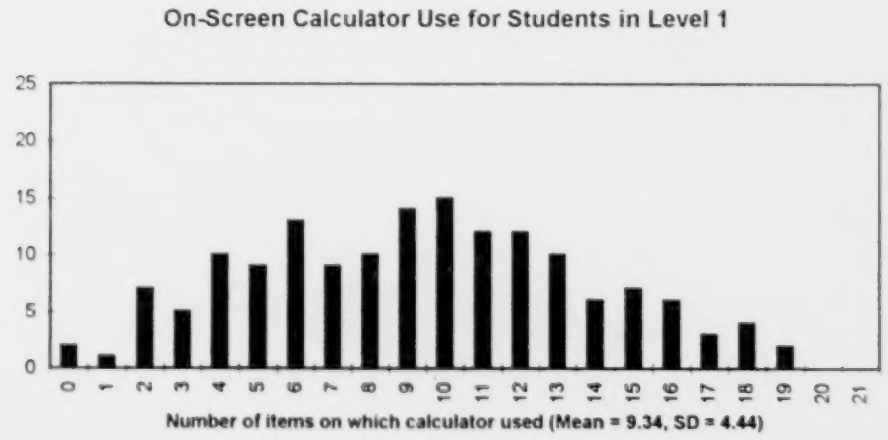
On-Screen Calculator Use for Students in Level 1
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Number of items on which calculator used (Mean = 9.34, SD = 4.44)

FIGURE 1. On-screen calculator use by Level 1 students.

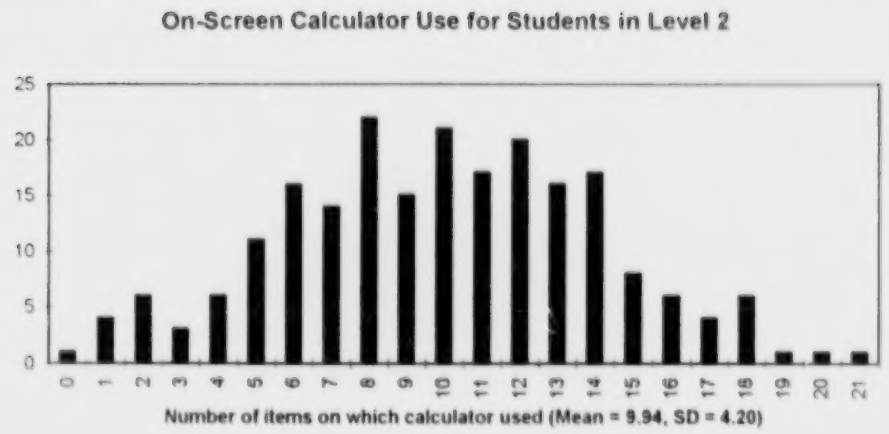
On-Screen Calculator Use for Students in Level 2
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Number of items on which calculator used (Mean = 9.94, SD = 4.20)

FIGURE 2. On-screen calculator use by Level 2 students.

Conclusion

Students who could use only an on-screen calculator appeared to do just as well on the computerized SAT-M as students who could use their own calculators. This finding appears to generalize across ability, gender, and ethnic groups, and across groups with varying degrees of calculator sophistication. Nevertheless, most students would strongly prefer to use their own calculators. There was substantial variation in how frequently students used a calculator with suggestive evidence that students who used a calculator for more than half of the questions might do better with their own calculators. If the SAT-M had stricter time limits, there might be an advantage to using an off-screen calculator. Any decision to change the current calculator policy would need to weigh the security benefits and perception of greater fairness of an on-screen calculator against the strong preference of students to use their own calculators.

Reference

Bridgeman, B., Harvey, A., & Braswell, J. (1995). Effects of calculator use on scores on a test of mathematical reasoning. Journal of Educational Measurement, 32, 323-340, 1995